

**No title available.**Patent Number: ☐ DE19610066

Publication date: 1997-09-18

Inventor(s): GRIMMBERGER MATHIAS (DE); JOURJINE ALEXANDER N DR (DE); KUNSTMANN NIELS DR  
RER NAT (DE); MUELLER MICHAEL (DE); WEBER FRANK (DE)

Applicant(s): SIEMENS NIXDORF ADVANCED TECHN (DE)

Requested  
Patent: ☐ WO9734251Application  
Number: DE19961010066 19960314Priority Number  
(s): DE19961010066 19960314IPC  
Classification: G06K9/62 ; G07C9/00 ; A61B5/00 ; A61B5/117EC  
Classification: G07C9/00C2D, G06K9/00FEquivalents: ☐ EP0886833 (WO9734251), B1

---

**Abstract**

---

In order to capture facial data, several pictures of a person are taken with a camera and checked for the presence of a face and the eye positions are determined. Features localised at grid points of a superposed grid, and determined by the positions of the eyes are extracted and summarised to form characteristics vectors. From the available sets of characteristics vectors, a set of reference characteristics vectors is calculated by averaging for each person. Using an optimisation process applied to a pre-existing data bank which assigns weightings to the reference characteristics vectors, the discriminance of the averaged characteristics is raised. The captured data are used as comparative data for the newly determined characteristics data of a single picture when determining or confirming a person's identity.

---

Data supplied from the esp@cenet database - I2

**THIS PAGE BLANK (USPTO)**

19 BUNDESREPUBLIK  
DEUTSCHLAND



DEUTSCHES  
PATENTAMT

12 Patentschrift  
10 DE 196 10 066 C 1

21 Aktenzeichen: 196 10 066.6-53  
22 Anmeldetag: 14. 3. 96  
43 Offenlegungstag: —  
46 Veröffentlichungstag  
der Patenterteilung: 18. 9. 97

51 Int. Cl.<sup>8</sup>:  
G 06 K 9/62  
G 07 C 9/00  
A 61 B 5/00  
A 61 B 5/117

DS

DE 196 10 066 C 1

Innerhalb von 3 Monaten nach Veröffentlichung der Erteilung kann Einspruch erhoben werden

73 Patentinhaber:

Siemens Nixdorf Advanced Technologies GmbH,  
01139 Dresden, DE

74 Vertreter:

Fuchs, F., Dr.-Ing., Pat.-Anw., 81541 München

72 Erfinder:

Grimmberger, Mathias, 01129 Dresden, DE; Jourjine,  
Alexander N., Dr., 01187 Dresden, DE; Kunstmann,  
Niels, Dr.rer.nat., 01139 Dresden, DE; Müller,  
Michael, 01217 Dresden, DE; Weber, Frank, 01099  
Dresden, DE

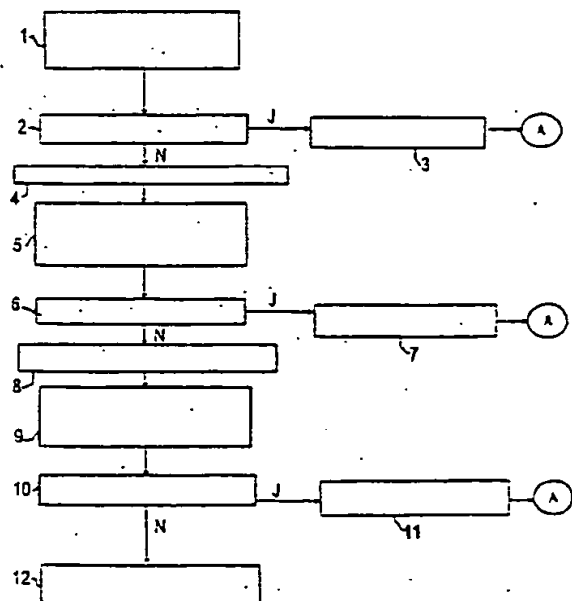
56 Für die Beurteilung der Patentfähigkeit  
in Betracht gezogene Druckschriften:

DE 44 06 020 C1 → 24) Fax 2  
WO 92 02 000 A1

Mustererkennung 1995, Informatik aktuell, Springer  
Verlag, S. 110-117;  
MOGHADDAM, B., PENTLAND, A.: »Face  
Recognition using View- Based and Modular  
Eigenspaces«, Automatic Systems for the  
Identification and Inspection of Humans, SPIE  
Vol. 2277, Juli 1994;  
RAO, R.P.N., BALLARD, D.H.: »Natural Basis  
Functions and Topographic Memory for Face  
Recognition«, Proc. of the Joint Conference on  
Artificial Intelligence, 1995, S. 10-17;  
ATICK, J.J., GRIFFIN, P.A., REDLICH, A.N.:  
»Statistical Approach to Shape from Shading:  
Reconstruction of 3D Face Surfaces from Single 2D  
Images« Computational Neuroscience Laboratory,  
The Rockefeller University, NY 10021-6399;  
KONEN, W., SCHULZE-KRÜGER, E.: »ZN Face: A  
system for access control using automated face  
recogni- tion« Intern. Workshop on Automatic Face-  
and Gesture-Recognition, Zurich, June 26-28, 1995,  
Proc., S. 18-23;

54 Verfahren zur Erfassung gesichtsbezogener Personendaten und deren Verwendung zur Identifizierung oder  
Verifikation von Personen

57 Zwecks Erfassung der gesichtsbezogenen Daten werden  
mehrere mit einer Kamera aufgenommene Bilder einer  
Person auf Vorliegen eines Gesichts überprüft und die  
Augenpositionen bestimmt. Sodann werden an durch die  
Augenpositionen bestimmten Gitterpunkten eines überlager-  
ten Gitters dort lokalisierte Merkmale extrahiert und aus den  
verfügbaren Merkmalsätzen ein Satz von Referenzmerkma-  
len je Person durch Mittelung berechnet. Durch einen  
Optimierungsprozeß gegenüber einer vorgegebenen Daten-  
bank, der den Referenzmerkmalen je nach Gitterpunkt  
unterschiedliche Gewichte zuordnet, wird die Diskriminanz  
der gemittelten Merkmale erhöht. Zur Identifizierung oder  
Verifikation einer Person dienen die erfaßten Daten als  
Vergleichsdaten für die dann jeweils neu ermittelten Merk-  
malsdaten.



DE 196 10 066 C 1

Auf dem Gebiet der Sicherheitstechnik ist es üblich, Personen an Hand von persönlichen Merkmalen zu identifizieren oder zu verifizieren, um ihnen Zugang zu sicherheitsrelevanten Einrichtungen zu gewähren.

Zu diesen persönlichen Merkmalen gehört auch das Gesicht, das zur Erfassung von gesichtsbezogenen Daten von einer Kamera aufgenommen wird. Die digitalen Bilddaten werden dann auf besondere Merkmale hin untersucht und ein entsprechender Graph gebildet, der in einer Datenbank abgespeichert wird und damit zum Vergleich mit einem jeweils neugebildeten Graphen zur Verfügung steht. Stimmt ein von einer Person neu abgeleiteter Graph mit einem abgespeicherten überein, erhält die Person den gewünschten Zugang.

Es ist bereit s eine Reihe von Gesichtserkennungssystemen bekannt, wobei die nachfolgende Numerierung sich auf das Literaturverzeichnis am Ende der Beschreibung bezieht.

Zu 1:

Der hierin dargestellte Ansatz verwendet zur Erkennung von Gesichtern deren Darstellung in einem Raum von Eigenfaces. Diese Darstellung ist die sogenannte Karhunen-Loève-Transformation (KLT). Sie erfordert eine große Menge von Trainingsmustern. Es seien diese für Bilder der Größe  $N \times N$  als  $N^2$ -dimensionale Vektoren  $x$  gegeben. Die Basisfunktionen der KLT erhält man dann durch Lösen des Eigenwertproblems

$$\Lambda = \Phi^T \Sigma \Phi$$

wobei  $\Sigma$  die Kovarianzmatrix der Datenvektoren,  $\Phi$  die Eigenvektor-Matrix von  $\Sigma$  und  $\Lambda$  die Diagonalmatrix der Eigenwerte ist. Man erhält die neue Darstellung von  $x$  als

$$\bar{y} = \Phi_M^T (\bar{x} - \bar{\bar{x}}),$$

wobei  $\bar{x}$  den Mittelwert aller  $\bar{x}$  bezeichnet.  $\Phi_M$  ist eine Submatrix von  $\Phi$ , die die führenden  $M$  Eigenvektoren, d. h. die Eigenvektoren zu den  $M$  größten Eigenwerten, enthält. In einer derartigen teilweisen KLT beträgt der Rekonstruktionsfehler für ein bestimmtes Muster  $x$  gerade

$$\varepsilon^2(\bar{x}) = \|\bar{x} - \bar{\bar{x}}\|^2 - \|\bar{y}\|^2$$

Diese Größe wird als "distance from feature space" bezeichnet und zur Lokalisierung des Gesichts innerhalb eines Bildes verwendet.

Solch ein Eigenface-basiertes System kann durch lokale Variation getäuscht werden. Weiterhin erfordert das System eine Normalisierungsstufe, die das Gesicht auf eine Standardgröße bringt. Da das Verfahren auch zur effizienten Kodierung von Gesichtern für die Videoübertragung verwendet wird, ist eine Erfassung mehrerer Bilder eines Gesichts mit nur einem Merkmalsatz prinzipiell nicht möglich.

Es werden daher zwei Erweiterungen vorgeschlagen, die diese Defizite teilweise beheben:

Einerseits wird das Schema der Eigenfaces auf beliebige andere Merkmale, wie z. B. Augen, Nase, Mund etc., des Gesichts ausgedehnt. Werden diese modularen Eigenvektor-Darstellungen mit den Eigenfaces kombiniert,

ergibt sich zwar eine bessere Erkennungsleistung, jedoch wird der bei jeder abbrechenden Eigenvektorentwicklung sich ergebende Nachteil, nämlich daß die erneute Berechnung der Eigenvektoren erforderlich wird, sobald neue Gesichter hinzukommen, nicht aufgehoben.

Andererseits wird versucht, mit Hilfe multipler Beobachter eine Erkennung zu realisieren, die von der Kopfrotation der Person unabhängig ist. In dem verwendeten formalen Rahmen bedeutet dies, daß für jeden dieser Beobachter die Koeffizienten der Entwicklung in dem zugehörigen Vektorraum bekannt sein müssen. Multiple Beobachter erfordern also multiple KLT und multiple Kodierungen. Für eine zu erfassende Person sind daher mehrere Datensätze zu speichern, was einen entsprechenden Aufwand an Speicherkapazität und Verarbeitungszeit erfordert.

Zu 2:

Das hierin beschriebene Verfahren zur Gesichtserkennung basiert im wesentlichen auf der Verwendung eines 2-Kamera-Systems zur Segmentierung. Als Merkmalsdetektoren werden Richtungsableitungen der zweidimensionalen Gaußfunktion verwendet. Der Startpunkt für die Definition und Konstruktion der Merkmalsdetektoren ist die bekannte Gaußfunktion

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Sie ist um  $x=0$  mit einer Varianz  $\sigma^2$  lokalisiert. Die einzelnen Detektoren sind Richtungsableitungen von  $G(x)G(y)$ , dem 2-dimensionalen Analogon von  $G(x)$ . Man bezeichnet sie üblicherweise mit  $G_n^{(k)}$ , wobei  $n$  die Ordnung der Ableitung und  $\theta_k = k\pi/(n+1)$  mit  $k=0, \dots, n$  ihre Richtung bezeichnet. Es werden Ableitungen der Ordnung 1 bis 3 verwendet.

Das System arbeitet als Identifikationssystem und die Assoziationsleistung wird von einem spärlich besetzten verteilten topographischen Speicher (SDM) erbracht. Zwar werden zum Training des SDM mehrere Bilder für jede Person verwendet, jedoch ist die erfaßte Information nicht nach Personen separierbar. Daher gibt es keine kompakte Darstellung des Aussehens einer einzelnen Person, wie es in einem Verifikationssystem erforderlich wäre.

Zu 3:

Das dargestellte System kann nur Folgen von Videobildern bei der Erkennung verarbeiten. Weiterhin werden als die Gesichter unterscheidende Merkmale die Koeffizienten einer Entwicklung der Oberfläche der Gesichter in 3-dimensionale Eigenköpfe verwendet. Somit zeigt das Verfahren die von den Eigenfaces bekannten Vor- und Nachteile.

Zu 4:

Bei diesem System wird der Vergleich zwischen dem zu erkennenden Bild und dem gespeicherten Bild durch einen flexiblen Abbildungsmechanismus realisiert. Zum Vergleich werden durch unspezifische Merkmalsdetektoren extrahierte lokale Merkmale des Gesichts verwendet. Die Gitterpositionen, an denen die Merkmale extrahiert werden, sind nicht explizit festgelegt, sondern variabel, wobei die Gitterverzerrung in der Vergleichsfunktion enthalten ist.

Die Positionierung der Gitterpunkte ergibt sich als Resultat eines globalen Optimierungsprozesses mit der Kostenfunktion

$$E = E_{\text{sim}} + \lambda E_{\text{top}}$$

wobei der erste Term  $E_{\text{sim}}$  die Ähnlichkeit extrahierter Merkmalsvektoren (Jets) an jeweils korrespondierenden Gitterpositionen und der zweite Term  $E_{\text{top}}$  die Gitterverzerrungskosten quantifiziert. Hier wird also ein bewegliches Gitter verwendet.

Dieses Verfahren hat daher folgende Nachteile:

Es funktioniert nicht bei komplexen Hintergründen, da das Gitter mit den Merkmalsdetektoren nicht ausschließlich das Gesicht überdeckt. Auch besteht innerhalb des formalen Rahmens keine Möglichkeit, mehrere Bilder mit nur einem Merkmalsatz zu erfassen. Weiterhin kann wegen der Vernachlässigung von Phaseninformation in der Konstruktion der Jets der Referenzdatensatz nicht zur Rekonstruktion des Gesichts verwendet werden.

Zu 5:

Das in der Patentschrift erwähnte Verfahren zur Erfassung und Wiedererkennung von Gesichtern basiert ebenso wie (1) auf dem Eigenface-Ansatz und ist somit ebenso nicht zur Erfassung mehrerer Bilder einer Person in einem kompakten Code geeignet. Als "Merkmale" gelten hier wie auch im modularen Ansatz von (1) tatsächliche, für den Menschen sichtbare und klar definierte Teile des Gesichts, wie z. B. Augen oder Augenbrauen. Außerdem wird ein dualer Code verwendet, der einerseits die absolute Position eines Merkmals ("off set") andererseits deren spezifisches Aussehen ("reference vertex") beschreibt. Um diese Merkmale zu finden, werden "intelligente" Mustererkennungstechniken (Stichwort "snake") angewandt.

Zu 6:

Das hier beschriebene Verfahren erweitert den in (4) beschriebenen Ansatz um eine Gewichtung der Gitterpositionen. Diese Gewichtung ist eindeutig für eine bestimmte Aufgabe und modifiziert die bisherige Ähnlichkeitsfunktion. Die Gewichte sind daher nicht Bestandteil des Referenzdatensatzes für eine erfaßte Person. Weiterhin erfolgt keine Spezifikation der Qualitätsfunktion  $Q$  zur Optimierung der Gewichte. Es ist lediglich angegeben, daß die Datenmenge  $T$ , die der Optimierung zugrunde liegt, nur jeweils Paare von Bildern verschiedener Personen enthält. Damit kann allenfalls ein konventionelles Erkennungsexperiment durchgeführt werden, und die Berücksichtigung mehrerer Bilder bei der Erfassung ist ausgeschlossen.

Die Gitterpositionen selbst ergeben sich vergleichbar dem Verfahren nach (4) als Resultat eines globalen Optimierungsprozesses, der das gesamte Gitter umfaßt. Die Lage eines einzelnen Gitterpunktes ist also nicht Resultat eines expliziten Detektionsprozesses und somit an die Existenz eines lokalen Merkmals gebunden, sondern ergibt sich aus dem kooperativen Verhalten des beweglichen Gitters.

Die Erfindung hat ein Verfahren zur Erfassung von gesichtsbezogenen Personendaten und deren Verwendung zur Identifizierung oder Verifikation in Zugangskontrollsystemen zum Gegenstand, das Merkmale der vorangehend genannten Systeme aufgreift und wesentlich erweitert, sowie eine schnelle und robuste Erkennung von menschlichen Gesichtern aus Standbildern unter natürlicher Beleuchtung ermöglicht.

Ein entsprechendes Verfahren ergibt sich aus den Merkmalen des Anspruchs 1.

Danach werden die gesichtsbezogenen Daten nicht von einem einzigen Bild, sondern von einer Reihe von Bildern der zu erfassenden Person abgeleitet, und das zu überlagernde Gitter orientiert sich mit seinen Gitter-

punkten an der Lage der Augen, die am einfachsten von den Gesichtsmerkmalen zu lokalisieren sind.

Anhand der extrahierten Merkmale je Gitterpunkt werden dann ein repräsentativer Merkmalsatz für alle Bilder sowie optimierte Gewichte für jeden Gitterpunkt berechnet, um die Gleichfehlerrate möglichst klein zu halten. Das bedeutet im Einzelnen:

Für jedes aufgenommene Bild  $i$  werden dabei  $d$  bestimmte Merkmalsdetektoren, z. B. solche wie in (2), an jedem Gitterpunkt  $g$  und für jede Auflösungsstufe  $k$  einer festen Menge von Auflösungsstufen des Bildes angewandt, so daß sich  $d$ -dimensionale Vektoren  $\vec{m}_g^{(k,i)}$  ergeben. Ein Merkmal ist hier eine Komponente eines solchen Vektors und daher im Gegensatz zu (5) die Antwort eines unspezifischen Merkmalsdetektors. Damit sind die Merkmale im Gesicht nicht klar lokalisiert, sondern man könnte sie statt dessen als delokalisiert bezeichnen.

Im nächsten Schritt werden jeweils mehrere Vektoren  $\vec{m}_g^{(k,i)}$  zu neuen Vektoren zusammengefaßt, die im folgenden als Merkmalsvektoren bezeichnet werden. Hierbei gibt es verschiedene Möglichkeiten: Beispielsweise kann die Zusammenfassung für jeden Gitterpunkt über mehrere Auflösungsstufen oder für jede Auflösungsstufe über mehrere Gitterpunkte erfolgen; denkbar ist auch eine Kombination dieser beiden Techniken. Die einfachste Möglichkeit ist, als Merkmalsvektoren die Vektoren  $\vec{m}_g^{(k,i)}$  selbst zu verwenden.

In jedem Falle existieren nach der Zusammenfassung Merkmalsvektoren  $\vec{n}_q$ ,  $q = 1, \dots, N_q$  wobei  $N_q$  die Zahl der Merkmalsvektoren bezeichnet. Vor der weiteren Verarbeitung sind die Merkmalsvektoren auf Länge 1 zu normieren; im folgenden wird daher angenommen, daß  $(\vec{n}_q)^2 = 1$  gilt.

In einem Mittelungsprozeß werden diese Sätze von Merkmalen aller aufgenommenen Bilder zu einem repräsentativen Referenzmerkmalssatz zusammengefaßt. Dieser repräsentative Referenzmerkmalssatz kann durch unterschiedliche Mittelungsschemata erzeugt werden. Betrachtet man nämlich  $\vec{n}_q$  als Satz normierter Merkmalsvektoren einer festen Person für die unterschiedlichen Trainingsbilder  $i = 1, \dots, N$  und für ein festes  $q$ , so erhält man (im Folgenden ist zur Vereinfachung der Notation der Index  $q$  weggelassen) einen repräsentativen Datensatz  $n = \sum_i \lambda_i \vec{n}_i$  durch die Minimierung einer Funktion  $V$ , die die Breite der Selbstähnlichkeit  $p_{ij} = \vec{n}_i \cdot \vec{n}_j$  aller Merkmalsvektoren quantifiziert, wobei die Normalisierungsnebenbedingung  $\vec{n}^2 = 1$  eingehalten werden muß. Im Fall  $V = \sum_i \lambda_i \vec{n}_i$  ergibt sich  $n \propto \sum_i \lambda_i \vec{n}_i$ , während sich für  $V = \sum_i (\vec{n}_i \cdot \vec{n})^2$  die Koeffizienten  $\lambda$  als Eigenvektor der Matrix  $P = (p_{ij})$  ergeben, der zum größten Eigenwert gehört. Diese Art Eigenvektor hat aber nichts mit einer KLT zu tun.

Alternativ kann derjenige Merkmalsatz unter den Merkmalsätzen der aufgenommenen Bilder als neuer repräsentativer Referenzmerkmalssatz bestimmt und verwendet werden, der den kleinsten Abstand, z. B. den kleinsten euklidischen Abstand, zu dem nach einer der vorhergehend beschriebenen Methoden ermittelt wurde.

Die Gesamtähnlichkeit eines Referenzmerkmalssatzes zu dem Merkmalsatz aus Bild  $i$  ergibt sich als gewichtete Summe der einzelnen Ähnlichkeiten

$$p_i^q = \vec{n}_i^q \cdot \vec{n}^q$$

zum jeweiligen Referenzmerkmalssatz  $\vec{n}^q$  in der Form

$$s(i) = \sum_q w_q p_i^q.$$

Die einzelnen Gewichte sind nicht negativ und ihre Summe beträgt 1.

Zusätzlich können auch die einzelnen Merkmale gewichtet werden. Dazu wird jedem Referenzmerkmalsvektor  $\vec{n}^i$  eine Diagonalmatrix  $C^i$  zugeordnet, deren Diagonalelemente die Gewichte für jede Komponente von  $\vec{n}^i$  und damit für jedes einzelne Merkmal darstellen. Die Gewichtung erfolgt dann durch Multiplikation der Matrix mit dem Referenzmerkmalsvektor. Auch der gewichtete Referenzmerkmalsvektor muß auf Länge 1 normiert sein, d. h. es muß  $(C_q \vec{n}^i)^2 = 1$  gelten. Die einzelnen Ähnlichkeiten  $p_i^q$  ergeben sich dann nach folgender Gleichung:

$$p_i^q = \vec{n}^i \cdot C_q \vec{n}^i.$$

Die Berechnung von personenabhängigen Gewichten, die ebenfalls Teil der später abgespeicherten gesichtsbezogenen Daten sind, ist auf die spezielle Aufgabe zugeschnitten und berücksichtigt Informationen aus den aufgenommenen Bildern und einer Hintergrunddatenbank von Merkmalsätzen, was im Folgenden beschrieben wird.

Das Ziel jeden Verifikations- oder Identifikationssystems ist, gleichzeitig die Rate der falschen Zurückweisungen (false rejection rate, FRR,  $\chi^{(1)}$ ) und die Rate der falschen Akzeptierungen (false acceptance rate, FAR,  $\chi^{(2)}$ ) möglichst gering zu halten. In unserem System sind beides Funktionen der Akzeptanzschwelle  $\lambda$  und definiert als

$$\chi^{(1)}(\lambda) = \frac{1}{N_A} \sum_{i=1}^{N_A} \theta \left( \lambda - \sum_{q=1}^{N_q} w_q p_i^q \right)$$

$$\chi^{(2)}(\lambda) = \frac{1}{N_B} \sum_{j=1}^{N_B} \theta \left( \sum_{q=1}^{N_q} w_q r_j^q - \lambda \right)$$

mit  $\theta(x)$  als Heavisidesche Stufenfunktion. Hier ist  $N_A$  die Anzahl der unterschiedlichen Bilder von Person A, die bei der Erfassung aufgenommen werden, während  $N_B$  die Anzahl der Merkmalsätze von Bildern anderer Personen in der Hintergrunddatenbank ist.  $p_i^q$  bezeichnet die Ähnlichkeit des Merkmalsvektors  $\vec{n}^i$  des Bildes i von Person A zum Referenzmerkmalsvektor  $\vec{n}^q$  und  $r_j^q$  den entsprechenden Wert für eine Person  $B \neq A$ .

Das oben genannte Ziel kann durch Bestimmung solcher personenabhängigen Gewichte erreicht werden, die eine geeignete Zielfunktion optimieren. Sinnvolle Zielfunktionen sind beispielsweise die Summe der beiden Fehlerraten (FAR und FRR) oder die Gleichfehler-rate (equal error rate, EER), der Funktionswert also, für dessen Argument FAR und FRR gleich sind. Eine andere Zielfunktion geht direkt aus von der Menge der Gesamthähnlichkeiten zwischen dem Referenzmerkmals-satz von Person A und den Bildern von A ( $p_i = \sum_q w_q p_i^q$ ) sowie von der Menge der Gesamthähnlichkeiten zwischen diesem Referenzmerkmalsatz und den Merkmalsätzen aus der Hintergrunddatenbank ( $r_j = \sum_q w_q r_j^q$ ) und quantifiziert das Ausmaß der Trennung dieser Mengen:

$$f(w) = \frac{\bar{p}_i - \bar{r}_j}{\sqrt{\text{Var}(p_i) + \text{Var}(r_j)}}$$

Hierbei bezeichnen  $\bar{p}_i$  und  $\bar{r}_j$  die jeweiligen Mittelwerte der Gesamthähnlichkeiten über alle i bzw. j sowie  $\text{Var}(p_i)$  und  $\text{Var}(r_j)$  die jeweiligen Varianzen. Alle drei genannten Zielfunktionen ergeben sich bei gegebenen p und r als Funktionen der Gewichte. Zusätzlich zu den Gewichten muß noch der optimale Schwellwert bestimmt werden:

Bei den beiden erstgenannten Zielfunktionen ergibt er sich bei der Optimierung der Gewichte, bei der dritten Zielfunktion kann er nach der Gewichtsoptimierung z. B. als Ort der Gleichfehlerrate gewählt werden.

Optimierungsverfahren mit Bestimmung von Gewichten sind bekannt, z. B. als stochastische Suche im Raum der möglichen Gewichte, siehe z. B. "Numerical Recipes in C" von Press et al, Cambridge University Press, 2. Auflage 1993.

Weiterbildungen der Erfindung ergeben sich aus den Unteransprüchen, die sich u. a. auf die unterschiedlichen Möglichkeiten für die Berechnung repräsentativer Referenzmerkmalsätze und auf die zusätzliche Verwendung der Nasenlage für die Bestimmung der Gitterpunktpositionen beziehen.

Einzelheiten der Erfindung seien nachfolgend an Hand von in der Zeichnung dargestellten Verfahrensabläufen näher erläutert. Im einzelnen zeigen

Fig. 1A das Flußdiagramm eines Bilderfassungsprozesses,

Fig. 1B das Flußdiagramm eines anschließenden Optimierungsprozesses und

Fig. 2 das Flußdiagramm für eine Verifikation.

In diesen Flußdiagrammen sind die einzelnen Verfahrensschritte fortlaufend numeriert und die nachfolgende Erläuterung ist entsprechend dieser Numerierung gegliedert.

Mit Bezug auf Fig. 1A sind das folgende Verfahrensschritte:

1. Innerhalb der durch das System vorgegebenen Zeit von 10 Sekunden werden 8 Bilder aufgenommen. Für jedes zweite Bild werden dabei sofort die Augen lokalisiert. Bei den übrigen Bildern wird nur das Gesicht lokalisiert.

2. Die extrahierten Augenpositionen dienen dann zur Beurteilung, ob der Erfassungsprozeß erfolgreich war. Ein Erfassungsprozeß ist erfolgreich, falls die Augen in mindestens 3 der 4 Bilder gefunden wurden und gleichzeitig die Varianz der extrahierten Augenpositionen eine Schwelle überschreitet.

3. Sollte dies der Fall sein, ist die Erfassung beendet.

4. Im anderen Fall werden alle Bilder gespeichert, außer denjenigen, bei denen keine Augen lokalisiert werden konnten.

5. Wiederum werden 8 Bilder aufgenommen, von denen 4 der Augenfindung unterworfen werden. Wie bei der ersten Gruppe von 8 Bildern wird die Varianz der Augenpositionen berechnet.

6. Diese gilt zusammen mit der Anzahl der Bilder, bei denen der Augenfinder erfolgreich war, als Kriterium, ob die Erfassung erfolgreich war.

7. Im positiven Fall ist die Erfassung beendet.

8. Sollte dies nicht der Fall sein, so werden die Bil-

der mit lokalisierten Augen gespeichert.

9. siehe 5.

10. siehe 6.

11. siehe 7.

12. Falls die in 2. erwähnten Bedingungen nach keinem der drei Erfassungsschritte erfüllt wurde, gilt die Erfassung als nicht erfolgreich, und der Benutzer wird entsprechend informiert.

Bei erfolgreich abgeschlossener Erfassung der vorgegebenen Anzahl von Bildern wird auf den Optimierungsprozeß von Fig. 1B übergeleitet:

13. Bei den abgespeicherten Bildern, bei denen bisher keinerlei Verarbeitung durchgeführt wurde, wird nun versucht, das Gesicht und dann die Augen zu lokalisieren. Bilder, bei denen dieses nicht möglich ist, werden verworfen, und es werden die Augenpositionen zu den übrigen Bildern abgespeichert.

Bei allen danach gültigen Bildern werden die Position und die Skalierung des Gesichts bezüglich der Augenpositionen normalisiert, indem das jeweilige Bild rotiert und skaliert wird, bis die Augen innerhalb des Bildes auf einer horizontalen Linie liegen und einen vorgegebenen Abstand aufweisen. Für Bilder, bei denen die Augen gefunden wurden, werden auf verschiedenen Auflösungsstufen die Merkmalsätze bestimmt.

14. Auf jeder Auflösungsstufe werden an jedem Gitterpunkt die Referenzmerkmale durch Mittelung über die Merkmale der einzelnen Bilder bestimmt.

15. Zu Beginn des Optimierungsprozesses werden die relativen Gewichte der Gitterpunkte mit dem Kehrwert der Anzahl der Gitterpunkte initialisiert. Der Zähler  $i$  für die Optimierungsschritte wird auf 0 gesetzt.

16. Für alle Auflösungsstufen wird die EER (equal error rate, Kreuzungspunkt der Kurven für falsche Akzeptanz und falsche Zurückweisung) des Referenzmerkmalssatzes berechnet und unter EER( $i$ ) abgespeichert. Die Datenbank für diese Berechnung besteht aus den Merkmalsätzen aller bei der Erfassung akzeptierten Bilder und einer Datenbank von Merkmalen für andere Personen.

17. Wahl einer zufälligen Kombinationen von Gewichten für die Gitterpunkte. Der Schleifenzähler  $i$  wird um 1 erhöht.

18. Sollte die maximale Anzahl von Schleifendurchläufen überschritten sein, ist die Optimierung beendet und es wird mit Schritt 22 fortgesetzt.

19. Die neuen Merkmale mit geringster EER, sowie die Schwellen für jede Auflösungsstufe werden gespeichert.

20. Für die zufällige Wahl der Gewichte werden die EERs und die Schwellenwerte auf allen Auflösungsstufen berechnet. Die jeweiligen Resultate werden als EER( $i$ ) abgespeichert.

21. Falls die neuen Fehlerraten geringer sind als die alten als beste gespeichert, so wird mit Schritt 19 fortgesetzt. Im anderen Falle folgt Schritt 17.

22. Nach Beendigung der Optimierung werden die Gewichte der Gitterpunkte für die niedrigste EER gespeichert. Ebenfalls gespeichert wird die Auflösungsstufe, auf der diese niedrigste Rate auftrat, sowie der zugehörige personenabhängige Schwellenwert.

Das Verfahren für die Verifikation einer Person umfaßt gemäß dem Flußdiagramm von Fig. 2 folgende Schritte:

31. Die PIN (personal identification number, persönliche Identifikationsnummer) der Person wird eingelesen.

32. Es wird kontrolliert, ob es sich um eine gültige Benutzernummer handelt.

33. Falls die Nummer dem System nicht bekannt ist, wird die Person zurückgewiesen.

34. Es wird kontrolliert, ob die eingegebenen Nummer zu einer autorisierten Person gehört.

35. Sollte das nicht der Fall sein, wird die Person abgewiesen.

36. Die Kamera nimmt ein Bild der Person auf, das über einen Framegrabber an den Computer weitergegeben wird.

37. Im aufgenommenen Bild wird versucht, ein Gesicht zu lokalisieren.

38. Die Ausgabe der Gesichtslokalisierung wird bewertet.

39. Wenn kein Gesicht gefunden wurde, wird die Person abgewiesen.

40. Der Augenfinder versucht, innerhalb des Bildes Augen zu finden.

41. Die Ausgabe der Augenfindung wird bewertet.

42. Falls keine Augen gefunden wurden, wird die Person abgewiesen.

43. Die Position und die Skalierung des Gesichts werden bezüglich der Augenpositionen normalisiert. Nach der Normalisierung liegen die Augen innerhalb des Bildes auf einer horizontalen Linie und haben einen festen, vorgegebenen Abstand. Dazu wird das Bild rotiert und skaliert.

44. Die Merkmale werden auf der Auflösungsstufe extrahiert, die für die entsprechende Person während der Erfassung als optimal ermittelt wurde.

45. Zwischen dem Satz der Referenzmerkmale und den aus dem aktuellen Bild extrahierten Merkmalen wird die Ähnlichkeit berechnet. Dabei werden die als optimal bestimmten Gewichte der Gitterpunkte berücksichtigt.

46. Die ermittelte Ähnlichkeit wird mit der personenspezifischen Schwelle verglichen.

47. Falls die Ähnlichkeit unterhalb der erforderlichen Schwelle liegt, wird die Person zurückgewiesen.

48. Falls die Ähnlichkeit über der Schwelle liegt, wird die Person akzeptiert.

Analoges gilt für die Identifizierung des Gesichts einer unbekannten Person mit Bezug darauf, ob die Person in der Datenbank für erfaßte Personen enthalten ist oder nicht.

#### Literaturverzeichnis:

1. B. Moghaddam und A. Pentland, Face Recognition using View-Based and Modular Eigenspaces, Automatic Systems for the Identification and Inspection of Humans, SPIE Vol. 2277, Juli 1994;
2. R. P. N. Rao und D. H. Ballard: "Natural Basis Functions and Topographic Memory for Face Recognition", Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 1995, pp. 10-17;
3. J. J. Atick, P. A. Griffin, A. N. Redlich: "Statistical

- Approach to Shape from Shading: Reconstruction of 3D Face Surfaces from Single 2D Images", Neural Computation 8, pp. 1321—1340, 1996;
4. W. Konen, E. Schulze Krüger: "ZN-Face: A system for access control using automated face recognition", International Workshop on automatic Face- and Gesture-Recognition, Zürich, June 26.—28, 1995, pp. 18—23. Dazu auch DE 44 06 020 C1;
5. PCT-Anmeldung Wo 92/02000 A1;
6. Mustererkennung 1995, Informatik aktuell, Springer Verlag, S. 110—117.

### Patentansprüche

1. Verfahren zur Erfassung gesichtsbezogener Personendaten, gekennzeichnet durch folgende Verfahrensschritte:
- Aufnahme mehrerer Bilder von der zu erfassenden Person,
  - Lokalisieren des Gesichtes in den aufgenommenen Bildern,
  - Lokalisieren der Augen im lokalisierten Gesicht,
  - Normalisieren des Gesichtes durch Rotieren und Skalieren des jeweiligen Bildes, bis die Augen innerhalb des Bildes auf einer horizontalen Linie mit vorgegebenem Abstand liegen,
  - Extraktion von Merkmalen durch Merkmalsdetektoren mit Bezug auf Gitterpunkte eines Gitters, wobei die Positionen der Gitterpunkte durch die Lage der Augen bestimmt sind, mit Zusammenfassung der Merkmale zu einem Merkmalsatz für das jeweilige Bild,
  - Berechnung eines repräsentativen Referenzmerkmalsatzes, der die Merkmalsätze aller aufgenommenen Bilder auf kompakte Weise zusammenfaßt,
  - Optimierung von personenabhängigen Gewichten für jeden Referenzmerkmalsatz zwecks Optimierung einer Zielfunktion, die eine Funktion dieser Gewichte sowie der Ähnlichkeitswerte zwischen dem Referenzmerkmalsatz und denjenigen Merkmalsätzen darstellt, die aus den aufgenommenen Bildern derselben Person und den Bildern anderer Personen aus einer Hintergrunddatenbank gewonnen wurden, mit Bestimmung einer optimalen Akzeptanzschwelle,
  - Abspeicherung der so ermittelten gesichtsbezogenen Daten der erfaßten Person.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Berechnung eines repräsentativen Referenzmerkmalsatzes mittels Durchschnittsbildung und Normalisierung mit Bezug auf alle Bilder der zu erfassenden Person erfolgt.
3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die Berechnung eines repräsentativen Referenzmerkmalsatzes mittels Ableitung des Eigenvektors zum größten Eigenwert der Ähnlichkeitsmatrix erfolgt.
4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß die Positionen der Gitterpunkte zusätzlich durch die Lage der Nase bestimmt werden.
5. Verfahren nach einem der Ansprüche 1 bis 4, gekennzeichnet durch die Verwendung der erfaßten und abgespeicherten gesichtsbezogenen Personendaten zur Personenidentifizierung, indem je-

- weils neu erfaßte Merkmalsdaten mit den gespeicherten Referenzdaten verglichen werden.
6. Verfahren nach einem der Ansprüche 1 bis 4, gekennzeichnet durch die Verwendung der erfaßten und abgespeicherten gesichtsbezogenen Personendaten zur Personenverifizierung, indem jeweils neu erfaßte Merkmalsdaten mit den durch eine zusätzliche persönliche Kennung ausgezeichneten gespeicherten Referenzdaten verglichen werden.

Hierzu 3 Seite(n) Zeichnungen



- Leerseite -

FIG 1a

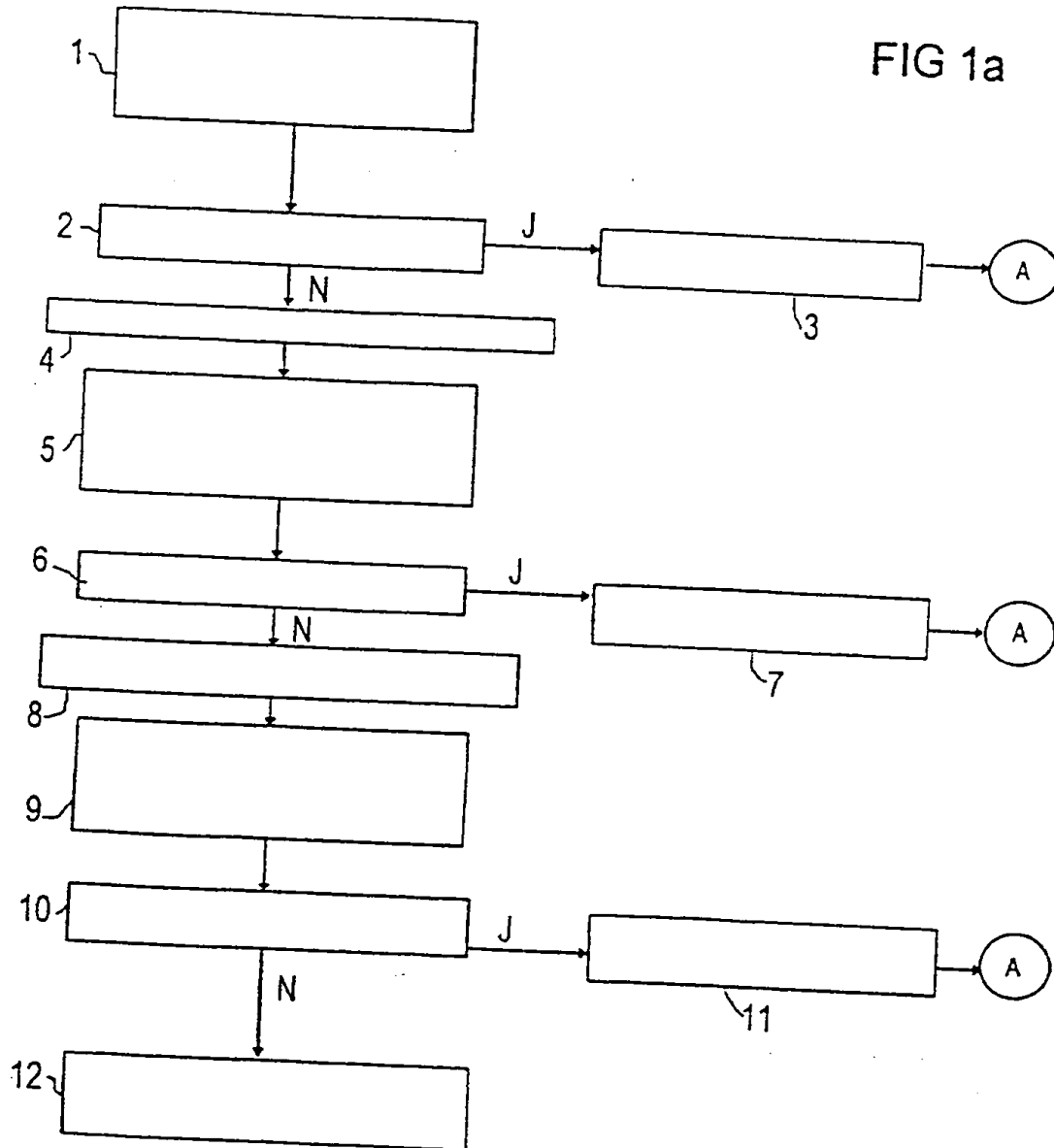


FIG 1b

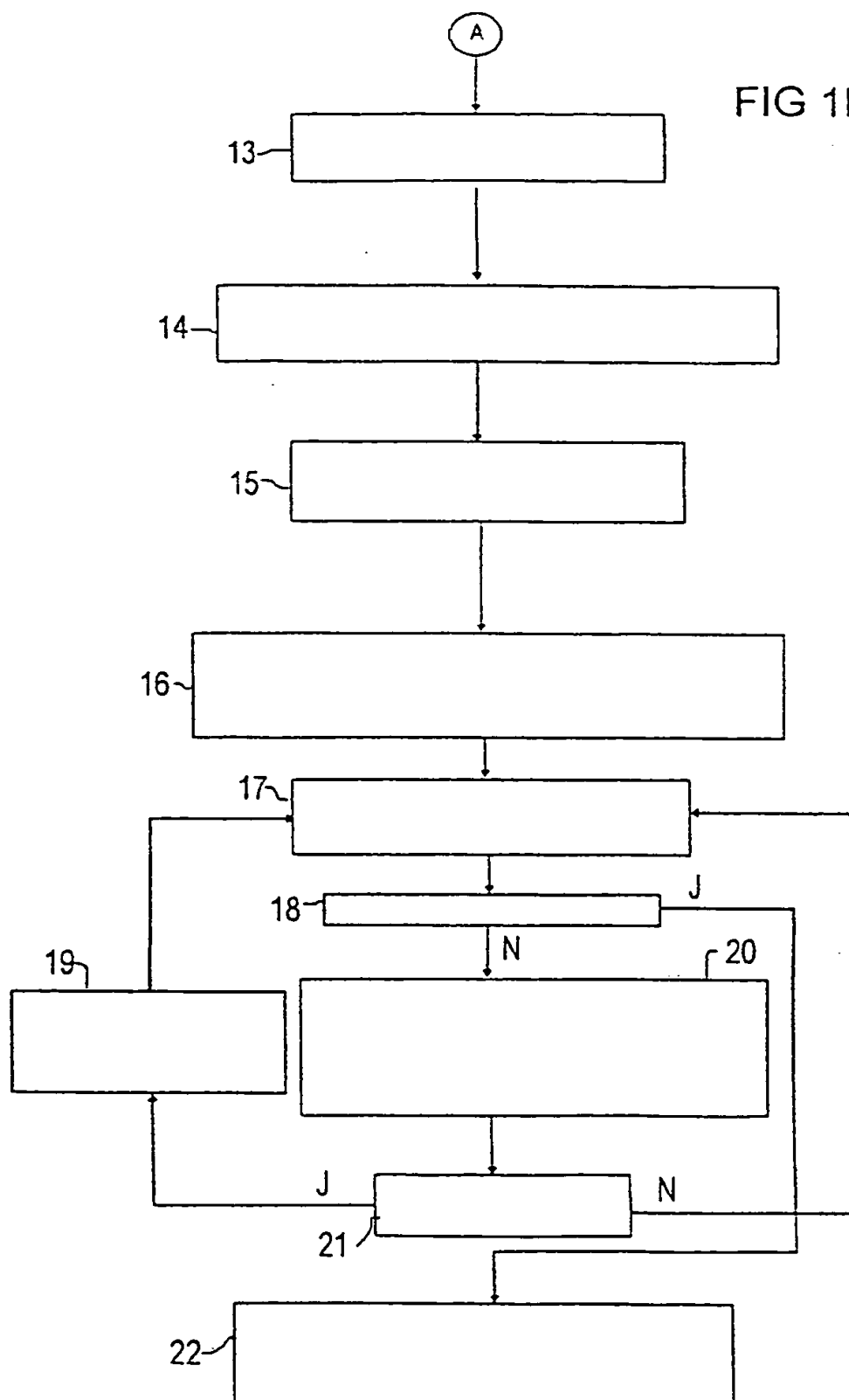


FIG 2

